Desaturase Enzymes

The invention relates to transgenic cells transformed with nucleic acid molecules which encode enzymes with desaturase activity and the use of these cells and enzymes in biocatalysis.

Desaturases are enzymes involved in the synthesis of long chain polyunsaturated fatty acids (PUFAs). PUFAs are fatty acids (FAs) which are essential to the normal functioning of a cell and their nutritional properties are well known. An example of a PUFA is docosahexanoic acid (DHA). DHA is a n-3 fatty acid that can be obtained directly from the diet or derived from metabolism of dietary linoleic and α-linolenic acid. The n-3 fatty acids are associated with health promoting properties. For example n-3 fatty acids have been described as anti-inflammatory, antithrombotic, antiarrhythmic, hypolipidemic and vasodilatory. As such, the role of DHA in the prevention and/or treatment of diseases such as coronary heart disease, hypertension, type II diabetes, ocular diseases, arthritis, cystic fibrosis and schizophrenia has been the focus of a great deal of medical research.

The production of PUFAs involves a consecutive series of desaturations and elongations of the fatty acyl chain to generate arachidonic acid (20:4Δ5,8,11,14) and docosahexaenoic acid (22:6Δ4,7,10,13,16,19). Several desaturases involved in this metabolic process have been isolated from marine microalgae, including *Phaeodactylum tricornutum* [5], *Euglena gracilis* [6] and *Pavlova lutheri* [7]. These membrane-bound desaturases are specific with respect to both chain length of the substrate and the double bond positions on the fatty acid. They belong to the class known as front-end fatty acid desaturases due to the fact that they introduce double bonds between the carboxy-group and pre-existing bond(s) of the fatty acid [1]. These desaturases contain a cytochrome *b5* domain at their N-terminus and three histidine motifs that are important for catalytic activity [10].

Desaturase enzymes and the genes which encode them are known in the art. For example, WO03/064596 describes, amongst other things, transgenic cells transformed with omega 3 and delta 12 desaturase nucleic acid molecules and the use of these cells in the production of fatty acids. In particular the use of the omega 3 desaturase in the conversion of arachidonic acid to eicosapentaenoic acid and the use of the delta 12 desaturase in the conversion of oleic acid to linoleic acid. WO03/099216 also describes fungal desaturases and in particular transgenic plants modified to express fungal delta 15 desaturase enzymes.

Furthermore, US2003/0157144 and US2003/0167525 disclose delta 5 and delta 6 desaturase genes in the conversion of dihomoylinolenic acid to arachidonic acid and linoleic acid to γ-linolenic acid respectively. Moreover, US2003/134400 discloses delta 4 desaturase genes which are involved in the conversion of adrenic acid to ω6-docosapentaenoic acid and in the conversion of ω3- docosapentaenoic acid to docosahexaenoic acid. These rare fatty acids are used in pharmacutical and cosmetic compositions and can be essential nutritional fatty acids.

Besides the common FAs 16:0, 16:1Δ9, 18:0 and 18:1Δ9 found in most living organisms, trace amounts of more unusual fatty acids can be found in a wide range of species. For instance, presence of 16:1Δ11 has been reported in several species of *Pavlova*, in the Eustigmatophyte *Nannochloropsis oculata*, and in the diatoms *Phaeodactylum tricornutum* and *Thalassiosira pseudonana* [11,12,13]. This FA accounted for a very small portion of the total FAs in these microalgae, and its specific role in the algal cells is unknown. However, this FA is a very important precursor in the synthesis of sex pheromones in insects. Sex pheromones are species-specific blends of unsaturated fatty acid (UFA) derivatives that differ in terminal functional group and in the number, position and configuration (Z or E) of the double bond(s), which are produced by various acyl-CoA desaturases [14,15]. Simple monoene Δ11 UFAs are the most prevalent precursors in the formation of major sex pheromome components in the modern *Lepidoptera* [16,17]. For instance, in the corn earworm *Helicoverpa zea*, which produces a pheromone mixture of Z11-16:Ald and

Z9-16:Ald in a 30:1 ratio, the most abundant desaturase-encoding transcript is *HzeaLPAQ* (also called *HzPGDs1*) which encodes a Δ11-desaturase that does not possess a cytochrome *b5* extension, and therefore requires free cytochrome *b5* for activity. Many acyl-CoA Δ11-desaturases with different specificities have been isolated from insects [14,15], but none from other species. We describe the first characterisation of a cytochrome *b5* desaturase exhibiting Δ11-desaturase activity.

According to an aspect of the invention there is provided a transgenic cell comprising a nucleic acid molecule which comprises a nucleic acid sequence which nucleic acid molecule consists of the sequences as represented in Figures 5a, 5b, 6a, 6c, 7a, 8a, 8b, 9a, 10a, 11a, 11b, 11d or nucleic acid molecules which hybridise to these sequences, wherein said nucleic acid molecules encode a polypeptide which has desaturase activity.

In a preferred embodiment of the invention said hybridisation conditions are stringent hybridisation conditions.

In a preferred embodiment of the invention said nucleic acid molecule comprises a nucleic acid sequence which has at least 30% homology to the nucleic acid sequence represented in Figures 5a, 5b, 6a, 6c 7a, 8a, 8b, 9a, 10a, 11a, 11b, 11d. Preferably said homology is at least 40%, 50%, 60%, 70%, 80%, 90%, or at least 99% identity with the nucleic acid sequence represented in Figures 5a, 5b, 6a, 7a, 8a, 8b, 9a, 10a, 11a, 11b and which encodes a polypeptide which has desaturase activity.

The sequence of desaturase nucleic acids may be modified to produce variant enzymes with enhanced expression in cells. For example, the addition of a codon that encodes an alanine amino acid may facilitate recombinant expression in microbial systems e.g. yeast. These modifications may not be required in all expression systems but is sometimes desirable.

In a preferred embodiment of the invention said nucleic acid molecule comprises the nucleic acid sequence as represented in Figures 5a, 5b, 6a, 6c 7a, 8a, 8b, 9a, 10a, 11a, 11b, 11d. Preferably said nucleic acid molecule consists of the nucleic acid sequence as represented in Figures 5a, 5b, 6a, 7a, 8a, 8b, 9a, 10a, 11a, 11b.

In a further preferred embodiment of the invention said cell over-expresses said desaturase encoded by said nucleic acid molecule.

In a preferred embodiment of the invention said over-expression is at least 2-fold higher when compared to a non-transformed reference cell of the same species.

Preferably said over-expression is: at least 3-fold, 4-fold, 5-fold, 6-fold, 7-fold, 8-fold, 9-fold, or at least 10-fold when compared to a non-transformed reference cell of the same species.

In a preferred embodiment of the invention said nucleic acid molecule is a cDNA.

In yet a further preferred embodiment of the invention said nucleic acid molecule is a genomic DNA.

In a preferred embodiment of the invention said transgenic cell is transfected with a nucleic acid molecule comprising a nucleic acid sequence as represented by Figure 10a and which encodes a desaturase polypeptide wherein said polypeptide has Δ11-desaturase activity, or a nucleic acid molecule which hybridises to the nucleic acid molecule in Figure 10a and encodes a polypeptide with Δ11-desaturase activity.

In an alternative preferred embodiment of the invention said transgenic cell is transfected with a nucleic acid molecule comprising a nucleic acid sequence as represented by Figure 8a and which encodes a desaturase polypeptide wherein said polypeptide has Δ6-desaturase activity, or a nucleic acid molecule which hybridises

to the nucleic acid molecule in Figure 8a and encodes a polypeptide with Δ6-desaturase activity.

In a preferred embodiment of the invention said transgenic cell is a eukaryotic cell.

In an alternative preferred embodiment of the invention said cell is a prokaryotic cell.

In a further preferred embodiment of the invention said eukaryotic cell is a plant cell.

Plants which include a plant cell according to the invention are also provided as are seeds produced by said plants.

In a preferred embodiment of the invention said plant is selected from: corn (*Zea mays*), canola (*Brassica napus*, *Brassica rapa* ssp.), flax (*Linum usitatissimum*), alfalfa (*Medicago sativa*), rice (*Oryza sativa*), rye (*Secale cerale*), sorghum (*Sorghum bicolor*, *Sorghum vulgare*), sunflower (*Helianthus annus*), wheat (*Tritium aestivum*), soybean (*Glycine max*), tobacco (*Nicotiana tabacum*), potato (*Solanum tuberosum*), peanuts (*Arachis hypogaea*), cotton (*Gossypium hirsutum*), sweet potato (*Iopmoea batatus*), cassava (*Manihot esculenta*), coffee (Cofea spp.), coconut (*Cocos nucifera*), pineapple (*Anana comosus*), citris tree (Citrus spp.) cocoa (*Theobroma cacao*), tea (*Camellia senensis*), banana (*Musa* spp.), avacado (*Persea americana*), fig (*Ficus casica*), guava (*Psidium guajava*), mango (*Mangifer indica*), olive (*Olea europaea*), papaya (*Carica papaya*), cashew (*Anacardium occidentale*), macadamia (*Macadamia intergrifolia*), almond (*Prunus amygdalus*), sugar beets (*Beta vulgaris*), oats, barley, vegetables and ornamentals.

Preferably, plants of the present invention are crop plants (for example, cereals and pulses, maize, wheat, potatoes, tapioca, rice, sorghum, millet, cassava, barley, pea), and other root, tuber or seed crops. Important seed crops are oil-seed rape, sugar beet, maize, sunflower, soybean,sorghum, and flax (linseed). Horticultural plants to which the present invention may be applied may include lettuce, endive, and

vegetable brassicas including cabbage, broccoli, and cauliflower. The present invention may be applied in tobacco, cucurbits, carrot, strawberry, sunflower, tomato, pepper.

Grain plants that provide seeds of interest include oil-seed plants and leguminous plants. Seeds of interest include grain seeds, such as corn, wheat, barley, rice, sorghum, rye, etc. Oil seed plants include cotton, soybean, safflower, sunflower, Brassica, maize, alfalfa, palm, coconut, etc. Leguminous plants include beans and peas. Beans include guar, locust bean, fenugreek, soybean, garden beans, cowpea, mungbean, lima bean, fava been, lentils, chickpea, etc.

According to a further aspect of the invention there is provided a seed comprising a plant cell according to the invention. Preferably said seed is from an oil seed plant.

According to a yet further aspect of the invention there is provided a reaction vessel comprising at least one polypeptide according to the invention, fatty acid substrates and co-factors wherein said vessel is adapted for the desaturation of said fatty acids substrates.

In a preferred embodiment of the invention said polypeptide is expressed by a cell according to the invention.

Preferably said cell is a eukaryotic cell, for example a yeast cell.

In an alternative preferred embodiment of the invention said cell is a prokaryotic cell.

According to a further aspect of the invention there is provided a method to desaturate a fatty acid substrate comprising the steps of:

i) providing a reaction vessel according to the invention; and

ii) growing said cells contained in said reaction vessel under conditions which allow the desaturation of at least one fatty acid substrate.

An embodiment of the invention will now be described by example only and with reference to the following tables and figures:

Table 1 illustrates the composition of major fatty acids in *T. pseudonana;*

Table 2 illustrates the major fatty acids of pYES and pYDESN yeast transformants with and without addition of exogenous saturated fatty acids;

Table 3 illustrates the Δ6 desaturase activity of TpDESI compared to that of an homologous *Phaeodactylum tricornutum* desaturase;

Figure 1 illustrates the predicted protein sequences with homology to front-end desaturases derived from the *T. pseudonana* draft genome. Sequence alignments of 12 putative *T. pseudonana* desaturases with other functionally characterised front-end desaturase enzymes identified three main blocks of homology that represent the functional domains of front-end acyl desaturases (A). The darker shaded box highlights the cytochrome *b5* haem-binding domain and shaded boxes indicate three histidine boxes. See Material and Methods for Genbank accession number and source species of the functionally characterised enzymes. A phylogenetic tree of nine *T. pseudonana* desaturases with other enzymes was constructed (B). By removing the regions containing gaps (ambiguous alignment region), a dataset was created from an alignment originally made with clustalX. The tree was constructed from the dataset using Phylip3.5c software package and bootstrap analyses were carried out with 1000 replicates. Only well supported nodes (over 70%) are indicated with bootstrap values. All branches are drawn to scale as indicated by the scale bar (=0.1 substitutions/site). TpDESN sequence is 477 amino acids long (C). The cytochrome *b5 haem*-binding domain is on a shaded background and the three histidine-boxes are framed;

Figure 2 illustrates RT-PCR expression analysis of *TpdesN*. Cells were harvested at different stages of growth for total RNA extraction and cDNA synthesis (A). PCR

was performed on cDNA derived from reverse transcribed RNA using *TpdesN* and *18s rRNA* specific primer pairs (B). PCR was carried out on undiluted (lane 1) and five-fold serial dilutions (lane 2-4) of each cDNA. The 18S rRNA gene was used as a control of cDNA synthesis. EE: early exponential phase, LE: late exponential phase, ES: early stationary phase;

Figure 3 illustrates GC analysis of FAMEs from yeast transformed with the empty plasmid pYES2 or the plasmid containing TpDESN. Invsc1 yeast strain transformed with either pYES2 (A) or pYDESN (B) were induced for three days at 20°C without supplementation before sampling for fatty acid analysis. I. S. internal standard (17:0). The experiment was repeated three times and results of a representative experiment are shown;

Figure 4 illustrates mass spectra of DMDS FAME adducts from pYDESN transformed yeast. Mass spectrum of the DMDS adduct of 16:1Δ9 FAME, present in all yeast samples (A). Mass spectrum of the DMDS adduct of 16:1Δ11 FAME, which was only found in yeast transformed with pYDESN (B). Picolinyl esters with spectra characteristic of 16:1Δ11 were also identified in these samples (data not shown); and

Figure 5a is the genomic nucleic acid sequence of the desaturase A from *T. pseudonana*; Figure 5b is the cDNA sequence desaturase A; Figure 5c amino acid sequence;

Figure 6a is the genomic nucleic acid sequence of desaturase B from *Thalassiosira pseudonana*; Figure 6b is the partial amino acid sequence; Figure 6c is the cDNA sequence of desaturase B; and Figure 6d is the amino acid sequence of said cDNA sequence;

Figure 7a is the nucleic acid sequence of desaturase E from *Thalassiosira psuedonana*; Figure 7b is the amino acid sequence;

Figure 8a is the nucleic acid sequence of desaturase I from *Thalassiosira pseudonana*; Figure 8b is the cDNA sequence; and Figure 8c is the amino acid sequence;

Figure 9a is the nucleic acid sequence of desaturase K from *Thalassiosira pseudonana;* Figure 9b is the amino acid sequence;

Figure 10a is the nucleic acid sequence of desaturase N from *Thalassiosira pseudonana*;

Figure 11a is the nucleic acid sequence of desaturase O from *Thalassiosira pseudonana*; Figure 11b is the cDNA sequence; Figure 11c is the amino acid sequence; Figure 11d is the nucleic acid sequence of desaturase O variant sequence from *Thalassiosira pseudonana*; and Figure 11e is the amino acid sequence of said variant desaturase O;

Figure 12A and 12B is a GC analysis of FAMEs from yeast expressing TpDESI with exogenous substrates 18:2Δ9,12 (A) and 18:3Δ9,12,15 (B). New FAs produced from endogenous and exogenous substrates are underlined; Figure 12C is a GC analysis of FAMEs from yeast transformed with a vector only control compared to yeast transformed with TpDESI;

Figure 13 is an illustration of fatty acid synthesis pathways; and

Figure 14 is a GC analysis of FAMEs from yeast expressing TpDESO.

Materials and Methods

Identification of putative Thalassiosira pseudonana desaturase-coding sequences and phylogenetic analysis with other functionally characterised desaturases

The draft genome of the diatom *T. pseudonana* has been sequenced to approximately nine times coverage by the whole genome shotgun method. Sequence data were produced by the US Department of Energy Joint Genome Institute (http://www.jgi.doe.gov/) and the raw sequence data were downloaded and installed on a local server. Batch tblastn searches were carried out using protein sequences of the following 13 known desaturases as query, including PlDES1 (AY332747, *Pavlova lutheri*), TFAD4 (AF489589, *Thraustochytrium sp.* ATCC 21685), TFAD5 (AF489588, *Thraustochytrium sp.* ATCC 21685), PtDEL5 (AY082392, *Phaeodactylum tricornutum*), PtDEL6 (AY082393, *Phaeodactylum tricornutum*), EgDEL8 (AF139720, *Euglena gracilis*), EgDEL4 (AY278558, *Euglena gracilis*), ZfDEL (AF309556, *Danio rerio*), BoDEL6 (U79010, *Borago officinalis*), HsDEL5 (AF084558, *Homo sapiens*), HsDEL6 (AF084559, *Homo sapiens*), CeDEL6 (AF031477, *Caenorhabditis elegans*) and CeDEL5 (AF078796, *Caenorhabditis elegans*).

All non-redundant sequences with an E value less than 0.001 were retrieved and assembled into contigs using the CAP3 sequence assembly program [18]. The contigs were translated into amino sequences in three frames in the orientation indicated by tblastn result. Putative desaturase gene models were constructed manually based on sequence homology and in frame GT-AG intron boundaries were identified.

Deduced amino acid sequences of all 12 putative desaturase sequences of *T. pseudonana* were aligned with the above 13 functionally characterised desaturases from other species, using ClustalX version 1.8 [19]. The alignment was then reconciled and further adjusted. Only nine near full-length *Thalassiosira* sequences were retained for further analyses.

A dataset of 250 conserved residue positions was used for construction of the phylogenetic tree. Distance analysis used the program protdist of the Phylip 3.5c package with a PAM250 substitution matrix and a tree was then built from the matrix

using fitch (Fitch-Margoliash method). Bootstrap analyses were carried out with 1000 replicates using the neighbour-joining algorithm.

Cultivation of *T. pseudonana*

T. pseudonana (CCAP 1085/12) was obtained from the Culture Collection of Algae and Protozoa (Dunstaffnage Marine Lab., Oban, PA34 4AD, Scotland, U.K.). The growth medium used was enriched artificial seawater medium (EASW), made up in 20 l batches as described previously [4]. The cultures were grown in one litre flasks at 15°C with 50 $\mu E\ m^{-2}\ s^{-1}$ constant illumination, and aeration provided by shaking the flasks at 150 rpm.

Cell density was monitored by counting cells with a haemocytometer. Nitrate concentration was determined periodically during the culture time by measuring the change of the medium absorbance at 220 nm [20].

RNA extraction, cDNA synthesis and RT-PCR analysis

Total RNA was extracted from frozen cells harvested at different stages of growth with an RNeasy plant mini kit (Qiagen). First strand cDNA was synthesised from three μg of DNAse treated RNA using a Prostar First-strand RT-PCR kit (Stratagene). PCR was performed using undiluted and five-fold dilutions of cDNAs as followed: the reactions were heated to 95 °C for 5 min followed by 35 cycles at 95 °C for 30 s, 50 °C or 65°C (for *18S rRNA* and *TpdesN* respectively) for 30 s and 72 °C for 2 min, then a single 72 °C for 10 min. As a marker for constitutive expression, the 18S rRNA gene was amplified with the primer TH18S5' (5'-GGTAACGAATTGTTAG-3') and TH18S3' (5'-GTCGGCATAGTTTATG-3'). *TpdesN* cDNA was amplified using primers DESNR2 (5'-GTGAGAGCACTAACCAAGCTT-3') and DESN2 (5'-CAATCAGTAGGCTTCGTC G-3'). Aliquots of PCR reaction were electrophoresed through a 1% agarose gel. Identity of the diagnostic fragment amplified with *TpdesN*

specific primers was verified by sequencing after cloning in the pGEM-T EasyVector (Promega).

Functional characterisation of TpDESI in yeast

The entire *TpdesI* coding region was amplified from *T. pseudonana* cDNA with primers DesINB 5'- GCGGGATCCACC**ATG***GCT*GGAAAAGGAGGAGAC-3' (ORF start codon is indicated by bold type; underlined sequence is a *BamH*I site; italic sequence is an added alanine codon, not present in the original sequence of *Pldes1*) and DesICE 5'-GCGAATTC**TTA**CATGGCAGGGAAATC-3' (ORF stop codon is indicated in bold type; underlined sequence is a *Eco*RI site). The Expand High Fidelity PCR system (Roche) was employed to minimise potential PCR errors. The amplified product was gel purified, restricted and cloned into the corresponding sites behind the galactose-inducible GAL1 promoter of pYES2 (Invitrogen) to yield the plasmid pYDES1. This vector was transformed into *S. cerevisiae* strain Invsc1 (Invitrogen) by a lithium acetate method, and transformants were selected on minimal medium plates lacking uracil.

For functional expression, cultures were grown at 25°C in the presence of 2% (w/v) raffinose and 1% (w/v) Tergitol NP-40 (Sigma). Expression of the transgene was induced when OD_{600nm} reached 0.2-0.3 by supplementing galactose to 2% (w/v). At that time, the appropriate fatty acids were added to a final concentration of 50 μM. Incubation was carried out at 25°C for three days.

Functional characterisation of TpDESN in yeast

Genomic DNA from *T. pseudonana* cells was extracted using the DNA isolation kit Puregene (Gentra Systems) and 100 ng was used to amplify the entire *TpdesN* coding region with primers DesNNB 5'-GCGGGATCCACC**ATG***GCT*GACTTTCTCTCCGGC-3' (ORF start codon is indicated by bold type; underlined sequence is a *BamH*I site; italic sequence is an

added alanine codon, not present in the original sequence of *TpdesN*) and DesNCE 5'-GCGAATTC**TCA**ATCAGTAGGCTTCGT-3' (ORF stop codon is indicated in bold type; underlined sequence is a *EcoR*I site). The Expand High Fidelity PCR system (Roche) was employed to minimise potential PCR errors. The amplified product was gel purified, restricted with *EcoR*I and *BamH*I and cloned into the corresponding sites behind the galactose-inducible GAL1 promoter of pYES2 (Invitrogen) to yield the plasmid pYDESN. The fidelity of the cloned PCR product was checked by sequencing. The vector pYDESN was then transformed into *S. cerevisiae* strain Invsc1 (Invitrogen) by a lithium acetate method, and transformants were selected on minimal medium plates lacking uracil.

For the feeding experiment with PUFAs, cultures were grown at 22°C in the presence of 2% (w/v) raffinose and 1% (w/v) Tergitol NP-40 (Sigma). Expression of the transgene was induced when OD_{600nm} reached 0.2-0.3 by supplementing galactose to 2% (w/v). At that time, the appropriate fatty acids were added to a final concentration of 50 μM. Incubation was carried out at 22°C for three days and then 15°C for another three days. For the feeding experiment with saturated fatty acids, a single Invsc1 colony transformed with pYES2 (empty plasmid, control) or pYDESN was inoculated in 10 ml of minimal media minus uracil containing 2% raffinose and grown overnight at 30°C with shaking (300 rpm). After 16-24 h, cells were collected by spinning at 4500 rpm for 10 min. After discarding the supernatant, the cell pellet was resuspended in the same medium mentioned above supplemented with 2% galactose and 1% tergitol, to obtain a cell density of 5×10^7 cells/ml. Fifteen ml of this cell suspension were added to a 100 ml-flask with or without addition of saturated fatty acids (as mentioned in the text) at 500 μM final concentration. Desaturase induction was then carried out at 20°C with shaking (300 rpm) for three days.

Fatty acid analysis

Microalgae or yeast cells were harvested by centrifugation. Total fatty acids were extracted and transmethylated as previously described [4]. Most FAMEs were

identified by comparison of retention times to a 37 FAME mix (Supelco). PUFA FAMEs were also identified by comparison to a sample of standard Menhaden oil (Supelco) transmethylated as per the samples.

Dimethyl disulphide (DMDS) adducts were used to determine the double bond position in identified and unidentified monounsaturated FAMEs. These were made by adding together 50 µl DMDS (Aldrich), 100-1000 ng FAMEs dissolved in 50 µl hexane, and 5 µl 50 mg ml^{-1} iodine in diethyl ether. This solution was heated at 40°C for 15 h and partitioned with 200 µl hexane and 100 µl 5% (w/v) sodium thiosulphate. The hexane phase was removed, dried under vacuum, reconstituted in 50 µl fresh hexane and used for GC-MS analysis. A Trace GC 2000 (ThermoQuest) fitted with a 30 m x 0.25 mm x 0.5 µm film thickness ZB−1 column (Phenomenex) was used to chromatograph 2 µl DMDS adducts injected at 250°C and a 50:1 split ratio with He as carrier gas at 0.6 ml min^{-1} in constant flow mode. The oven program was 120°C for 1min then to 340°C at 5°C min^{-1}. Mass spectra were obtained using a GCQ (ThermoQuest) mass spectrometer operating in full scan mode over 50-500 m/z. Picolinyl esters were also made from FAMEs to confirm their identities. These were obtained by adding 15 µl freshly prepared 2:1 (v/v) 3-(hydroxymethyl)-pyridine (Aldrich): potassium *tert* butoxide 1 M solution in tetrahydrofuran (Aldrich) to 50 µl FAMEs dissolved in dichloromethane. This solution was heated at 40°C for 30 min and partitioned with 200 µl hexane and 100 µl 2.5% (w/v) sodium hydrogen carbonate. The hexane phase was removed, dried under vacuum and reconstituted in 50 µl fresh hexane. Picolinyl esters were injected and separated by GC-MS using the same conditions as for DMDS adducts; Sperling P., Zahringer U. and Heinz E. (1998) A sphingolipid desaturase from higher plants. J. Biol. Chem. 273, 28590–28596; Sperling P., Libisch B., Zahringer U., Napier J.A. and Heinz E. (2001) Functional identification of a D8-sphingolipid desaturase from *Borago officinalis*. Arch. Biochem. Biophys. 388, 293–298; Whitney H.M., Michaelson, L.V., Sayanova, O., Pickett J.A. and Napier, J.A. (2003) Functional characterization of two two cytochrome b5-fusion desaturases from *Anemone leveillei*: The unexpected

identification of a fatty acid Δ6-desaturase. Planta 217, 983-992; each of which are incorporated by reference.

EXAMPLE 1

Identification and phylogenetic analysis of putative T. pseudonana desaturase sequences with other functionally characterised desaturases

Tblastn searches with 13 functionally characterised desaturases revealed 427 non-redundant raw sequences with E values less than 0.001. Twelve unique contigs were assembled after retrieving these sequences and gene models were constructed manually based on sequence homology. These 12 gene contigs were arbitrarily designated *TpdesA* to *TpdesL*. All 12 showed significant sequence similarity to query sequences with 9 containing near full length open reading frames compared to other known desaturases (Fig. 1A). Interestingly, the predicted amino acid sequence of all nine *T. pseudonana* desaturases have a characteristic fused cytochrome *b5 haem*-binding domain (HP[G/A]G) at their N-terminus and three histidine boxes (H[X]3-4H, H[X]2-3HH AND Q[X]2-3HH) with the replacement of the first histidine by glutamine in the third histidine box in all but two of the predicted proteins (TpDESA and TpDESB). These are common characteristics of a large subgroup of front-end acyl group desaturases [21]. These histidine-box motifs are critical for desaturase activity, most likely because they serve to coordinate the diiron-oxo component of the active site. Three remaining sequences (TpDESD, TpDESL and TpDESH) appear to be partial, covering only the C-terminal end of desaturases, but nevertheless they do contain a typical third histidine box of the above mentioned subgroup of desaturases (Fig. 1A).

In order to gain insight into the relationships of these *T. pseudonana* sequences to other functionally characterised desaturases and especially algal desaturases, we constructed an unrooted phylogenetic tree using a Fitch-Margoliash method with statistical confidence measured by bootstrap analysis (Fig. 1B). Relationships of four putative *T. pseudonana* desaturases are in well supported (>70% bootstrap value) subgroups with at least one functionally characterised desaturase from other species.

Both TpDESM and TpDESO grouped with PtDEL5, a Δ5-desaturase from another diatom, *P. tricornutum* [5], suggesting these two enzymes may also have a Δ5-desaturase activity. Similarly TpDESK is grouped with two Δ4-desaturases TFAD4 and EgDEL4 from *Thraustochytrium sp.* ATCC21685 [22] and *E. gracilis* respectively. TpDESI grouped with PtDEL6, a Δ6-desaturase from *P. tricornutum*. This indicates that TpDESK and TpDESI may have Δ4 and Δ6-desaturase activities respectively. However, as enzymes with different regioselectivities are also found in a well supported subgroup (EgDEL8, CeDEL5 and CeDEL6; Δ8, Δ5 and Δ6-desaturase respectively) and regioselectivity may even derive independently after a more recent duplication (CeDEL5 and CeDEL6) [23] predictions based on homology can be misleading and it is essential to functionally characterise each enzyme.

The remaining five *T. pseudonana* sequences fall into three separate subgroups (TpDESE; TpDESA and TpDESB; TpDESG and TpDESN) which do not group with any other known functional desaturases with high confidence. It is therefore possible that these proteins exhibit novel regioselectivity. The current study focussed on the characterisation of one of these proteins, TpDESN.

EXAMPLE 2

Temporal expression of TpDESN gene

RT-PCR analysis of *TpdesN* transcript was conducted at different stages of algal growth in order to establish if and when this gene is expressed. After RNA extraction and cDNA synthesis, *TpdesN* specific PCR products were amplified. PCR amplification of the 18S rDNA gene was performed as a control for the quantity of cDNA used during PCR reactions. Figure 2 shows that the diagnostic 519 bp cDNA amplification product expected for *TpdesN* was present at similar level at the different stages of cultivation of the microalga cells. Thus, *TpdesN* is transcriptionally active at a constitutive level during *Thalassiosira* growth, suggesting that it may encode a desaturase with a housekeeping function.

EXAMPLE 3

Functional characterisation of TpDESN in yeast

The putative desaturase sequence annotated *TpdesN* was contained on a genomic DNA contig of 2580 bp on which no introns was detected. To establish the function of the protein encoded by this gene, the full-length sequence was amplified from genomic DNA. An alanine codon containing a G as the first letter was added immediately downstream of the start codon of *TpdesN* to ensure optimal translation in yeast [24]. The *TpdesN* ORF is 1434 bp long, and encodes a 477 amino acid protein TpDESN (Fig. 1C), having a molecular weight of 53.8 kDa. Analysis of the secondary structure of TpDESN using SOSUI software (http://sosui.proteome.bio.tuat.ac.jp/sosuiframe0.html) [25] predicted four transmembrane regions (not shown). Alignment of TpDESN with functionally characterised desaturase sequences mentioned above indicated an overall identity of 25%, with the cytochrome *b5*-like domain and the three conserved histidine-rich motif areas showing greatest homology.

The primary sequence of TpDESN exhibited typical features of front-end desaturases involved in PUFA synthesis. In order to characterise the specificity of this protein, PUFAs (18:2Δ9,12; 20:2Δ11,14; 20:3Δ8,11,14; 22:4Δ5,8,11,14; 18:3Δ9,12,15; 20:3Δ11,14,17; 20:4Δ8,11,14,17; 22:5Δ7,10,13,16,19) where first fed to the host yeast transformed with pYDESN and the vector alone (pYES2) as a control. Unexpectedly, after six days of incubation, TpDESN did not desaturate any of the supplemented PUFA substrate. Furthermore, there did not appear to be any production of 18:2Δ9,12 from endogenous 18:1Δ9. However, a significant increase was observed for a peak eluting in the range of sixteen carbon monounsaturated FAMEs in the yeast transformed with pYDESN (Fig. 3). The position of the double bond in this product was determined by GC-MS analysis of FAMEs derived to DMDS adducts [26] and picolinyl esters. The DMDS adduct of 16:1Δ9 FAME yielded two major fragments at *m/z* 145 and 217 (Fig. 4A). Fragmentation of the

increased FAME peak found in unfed or fed yeast transformed with pYDESN produced two diagnostic fragments at *m/z* 117 and 245 (Fig. 4B). This fragmentation pattern was indicative of an Δ11 monounsaturated sixteen carbon FAME, 16:1Δ11, suggesting that TpDESN encoded a new Δ11-desaturase. Small amounts of this FA have also been measured in *Thalassiosira* cells (Table 1). To further substantiate these results, yeast transformed with pYDESN and the control empty vector, pYES2, were cultivated in medium supplemented with saturated FA (14:0; 16:0; 18:0) representing potential substrates for the synthesis of the monounsaturated product. Yeast fatty acid profiles were analysed after three days of incubation at 20°C. Results in Table 2 showed that a small amount of 16:1Δ11 (0.23% of total FAs) was detected in yeast transformed with pYES2, suggesting endogenous synthesis of this FA from 16:0. This FA accumulated at a higher level in both types of transformed yeast after feeding with 14:0, with values up to 5.84% in pYDESN transformants. A possible explanation for this increase in the pYES2 transformants is that the endogenous yeast Δ9-desaturase was able to use additional 14:0 to produce 14:1Δ9 that was subsequently elongated to 16:1Δ11. Moreover, it has been reported that wild type yeast cells cultivated in media supplemented with 14:1Δ9 synthesised 16:1 Δ11 by Elo1p-dependent carboxy terminal elongation [27]. After 18:0 supplementation, the percentage of 16:1Δ11, of about 6% total FAs, was similar to that observed after feeding with 16:0. Presence of extra 18:0 could lead to an inhibition of the 16:0 chain elongation system, which might allow more 16:0 to be available for Δ11-desaturation. On the other hand, 18:1Δ11 represents 1.2% of the total FAs in transgenic yeast. No variation in its proportion was monitored under the different conditions of incubation, even after supplementation with 18:0 in pYDESN transformants. This suggests that this FA originates from elongation of 16:1Δ9 rather than Δ11-desaturation of 18:0.

EXAMPLE 4

Functional characterisation of TpDESI in yeast

To establish the function of TpDESI, the full-length cDNA was expressed in the yeast Invsc1 under the control of an inducible galactose promoter. Potential

substrates of front-end desaturases (18:2Δ9,12; 18:3Δ9,12,15; 20:3Δ8,11,14; 20:4Δ8,11,14,17; 22:4Δ7,10,13,16; 22:5Δ7,10,13,16,19) were tested. Figure 12A and 12B show that after supplementation of the medium with 18:2$\Delta^{9,12}$ and 18:3$\Delta^{9,12,15}$ respectively, and after three days of incubation, yeast cells containing pYDESI had extra fatty acids. Extra peaks observed when cells were fed with 18:2Δ9,12 had a retention time identical to 16:2Δ6,9, 18:2Δ6,12 and 18:3Δ6,9,12 (Figure 12A). Extra peaks observed when cells were fed with 18:3Δ9,12,15 had a retention time identical to 16:2Δ6,9, 18:2Δ6,12 and 18:4Δ6,9,12,15 (Figure 12B). These results demonstrate that *TpdesI* encodes a Δ6-desaturase which can introduce double bond in exogenously fed 18:2Δ9,12 and 18:3Δ9,12,15 fatty acids, but also in endogenous 16:1Δ9 and 18:1Δ9 fatty acids. Percentages of conversion of these different substrates are given in Table 3.

Fatty acid profiling of marine microalgae had shown that *T. pseudonana* represents a good candidate to discover genes involved in the production and storage of PUFAs [4]. Analysis of the recently completed draft genome of this microalga revealed the presence of many candidate genes for elongase and desaturase activities most probably involved in catalysing different steps of the PUFA biosynthetic process. We have identified 12 possible desaturase genes, 9 of which there is sufficient sequence information to demonstrate that they exhibit typical features of front-end desaturases, *i.e.* a cytochrome *b5* domain in the N-terminus and three histidine clusters located at highly conserved regions. Phylogenetic analysis revealed that several of the genes are closely related to a number of previously characterised front-end desaturases involved in PUFA synthesis. However, the current work highlights the fact that desaturase function, in terms of regioselectivity, cannot solely be based on prediction from primary amino acid sequence homology.

The fatty acid profile of *T. pseudonana* cells is quite diverse (Table 1), with the health beneficial EPA (20:5Δ5,8,11,14,17) and DHA (22:6Δ4,7,10,13,16,19) accounting for a large proportion. However, the number of desaturase gene sequences found in the genome was higher than we expected based on the number of different

desaturation reactions required to produce the diversity of FA in this microalga. This suggested that non-obvious desaturation reactions might also occur in the *Thalassiossira* cells. As a first step to establishing function of the many putative desaturase sequences, we focused on the *TpdesN* contig due to the fact that the sequence was full-length and intronless. A temporal expression study showed that *TpdesN* was constitutively transcribed during algal cultivation. Expression of the *TpdesN* ORF in yeast supplemented with PUFAs as potential substrates for desaturation revealed no new products. There was also no evidence of activity with the endogenous 18:1Δ9 which excludes the possibility that TpDESN acts as a Δ12-desaturase. However, an increase in the peak area of a FAME eluting in the range of the sixteen carbon FAMEs was identified and GC-MS based analysis revealed this to be 16:1Δ11 fatty acid. Small amounts of this FA are also present in wild type yeast. However, quantitative comparison of FA levels in the empty vector pYES2 and pYDESN transformants showed that proportions of 16:1Δ11 increased in the presence of *TpdesN* in both unfed cells and cells that had been fed different saturated FA. No other changes in either peak area or new peaks were detected in pYDESN transformants, indicating that TpDESN is specifically involved in conversion of 16:0 to 16:1Δ11.

The presence of small amounts of 16:1Δ11 have previously been reported in many microalgae, including *T. pseudonana*. However, a function for this FA in algal cells has not been established. The low quantity observed in many marine microalgae suggests that it may act as an intermediate in an as yet unidentified biosynthetic pathway. In insect cells, 16:1Δ11 represents an important precursor for pheromone synthesis, where it is produced by an acyl-CoA Δ11-desaturase. Interestingly, the insect Δ11-desaturases do not possess a cytochrome *b5* domain in their N-terminal region. This represents a major primary structure difference compared with TpDESN. The cytochrome *b5* domain is not a determinant of the substrate specificity [28]. Alignment of the desaturase domain of TpDESN with the full sequence of insect Δ11-desaturases showed an identity of 20% (data not shown). In insect cells, Δ11-

desaturases are more or less specific depending on the origin of the sequence and well-documented reviews exist on this subject [14,15].

In conclusion therefore, although the TpDESN primary sequence is very similar to front-end desaturases, it should not be considered a member of this family of desaturases because it acts only on 16:0. Identification of such a novel enzyme expands the functional repertoire of the membrane-bound desaturases and it should provide useful comparative information for understanding phylogenetic relationships between these enzymes. One question that remains to be answered regards whether cytochrome *b5* was independently fused to desaturases that had already acquired their different specificities, or whether an ancestral fusion protein for proximal lipid modification duplicated and subsequently evolved into different desaturases. Studies of the primary structure of the different PUFA desaturases support the fact that enzyme conversion (i.e. change of specificity) can be achieved through a relatively few structural changes [29]. The high degree of homology between the many potential front-end desaturases identified in the genome of *T. pseudonana* support this notion. Given the FA profile of *T. pseudonana* cells and the complexity of the desaturase gene family it is likely that different genes will encode Δ4, Δ5 and Δ6 desaturases. It will now be very interesting to functionally characterise these remaining putative desaturase genes and study the relationship between regioselectivity, primary amino acid sequence and phylogenetic relationship. A crystal structure for these enzymes is still not available due to technical difficulties in obtaining sufficient quantities of purified membrane-bound protein. Molecular genetic approaches involving site-directed mutagenesis have provided new insight into structure-function relationships, including for example that residues in close proximity to the histidine motifs have been found to be involved in shifting the ratio of desaturation/hydroxylation activities [30]. Detailed comparative analyses and computer modeling of these diverse desaturases from *T. pseudonana* may further guide site-directed mutagenesis studies aimed at defining key residues controlling substrate specificity and regioselectivity of the introduced double bond.

References

[1] Napier, J.A., Sayanova, O., Sperling, P. and Heinz, E. (1999) Trends Plant. Sci. 4, 2-4.

[2] van de Loo, F.J., Fox, B.G. and Somerville, C.R. (1993) in: Unusual fatty acids (Moore, T.S., ed.) Lipid Metabolism in Plants, pp. 91-126.

[3] Vazhappilly, R. and Chen, F. (1998) J. Am. Oil Chem. Soc. 75, 393-397.

[4] Tonon, T., Harvey, D, Larson, T.R. and Graham, I.A. (2002) Phytochemistry 61, 15-24.

[5] Domergue, F., Lerchl, J., Zahringer, U. and Heinz, E. (2002) Eur. J. Biochem. 269, 4105-4113.

[6] Meyer, A., Cirpus, P., Ott, C., Schlecker, R., Zahringer, U. and Heinz, E. (2003) Biochemistry 42, 9779-9788.

[7] Tonon, T., Harvey, D., Larson, T.R. and Graham, I.A. (2003) FEBS Lett. 553, 440-444.

[8] Lopez Alonso, D., Garcia-Maroto, F., Rodriguez-Ruiz, J., Garrido, J.A. and Vilches, M.A. (2003) Biochem. Syst. Ecol. 31, 1111-1124.

[9] Pereira, S.L., Leonard, A.E. and Mukerji, P. (2003) Prost. Leuko. Essent. Fatty Acids 68, 97-106.

[10] Shanklin, J., Whittle, E. and Fox, D.C. (1994) Biochemistry 33, 12787-12794.

[11] Volkman, J.K., Dunstan, G.A., Jeffrey, S.W. and Kearney P.S. (1991) Phytochemistry 30, 1855-1859.

[12] Dunstan, G.A., Volkman, J.K., Barrett, S.M. and Garland, C.D. (1993) J. Appl. Phycol. 5, 71-83.

[13] Brown, M.R., Dunstan, G.A., Norwood, S.J. and Miller K.A. (1996) J. Phycol. 32, 64-73.

[14] Knipple, D.C., Rosenfield, C.-L., Nielsen, R., You, K.M. and Jeong, S.E. (2002) Genetics 162, 1737-1752.

[15] Roelofs, W.L., Liu, W., Hao, G., Jiao, H., Rooney, A.P. and Linn Jr, C.E. (2002) Proc. Natl. Acad. Sci. USA 99, 13621-13626.

[16] Roelofs, W.L. and Bjostad, L. (1984) Bioorg. Chem. 12, 279-298.

[17] Roelofs, W.L. and Wolf, W.A. (1988) J. Chem. Ecol. 14, 2019-2031.

[18] Huang, X. and Madan, A. (1999) Genome Res. 9, 868-877.

[19] Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F. and Higgins, D.G. (1997) Nucleic Acids Res. 25, 4876-4882.

[20] Collos, Y., Mornet, F., Sciandra, A., Waser, N., Larson, A. and Harrison, P.J. (1999) J. Appl. Phycol. 11, 179-184.

[21] Sperling, P., Ternes, P., Zank, T.K. and Heinz, E. (2003) Prost. Leuko. Essent. Fatty Acids 68, 73-95.

[22] Qiu, X., Hong, H. and MacKenzie, S.L. (2001) J. Biol. Chem. 276, 31561-31566.

[23] Michaelson, L.V., Napier, J.A., Lewis, M., Griffiths, G., Lazarus, C. and Stobart A. K. (1998) FEBS Lett. 439, 215-218.

[24] Kozak, M. (1987) Nucleis Acid Res. 15, 8125-8148.

[25] Hirokawa, T., Seah, B.C. and Mitahu, S. (1998) Bioinformatics 14, 378-379.

[26] Buser, H.R., Arn, H., Guerin P. and Rauscher, S. (1983) Anal. Chem. 55, 818-822.

[27] Schneiter, R., Tatzer, V., Gogg, G., Leitner, E. and Kohlwein, S.D. (2000) J. Bact. 182, 3655-3660.

[28] Libisch, B., Michaelson, L.V., Lewis, M.J., Shewry, P.R. and Napier, J.A. (2000) Biochem. Biophys. Res. Commun. 279, 779-85.

[29] Cahoon, E.B., Lindquist, Y., Schneider, G. and Shanklin, J. (1997) Proc. Natl. Acad. Sci. USA 94, 4872-4877.

[30] Broun, P., Shanklin, J., Whittle, E. and Somerville, C. (1998) Science 282, 1315-1317.